

# Exploring Two Novel Features for EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity

Nicolas Brodu <nicolas.brodu@univ-rennes1.fr>\*;  
Fabien Lotte <fprlotte@i2r.a-star.edu.sg>†;  
Anatole Lécuyer <anatole.lecuyer@irisa.fr>‡

August 2010

## Abstract

In this paper, we introduce two new features for the design of electroencephalography (EEG) based Brain-Computer Interfaces (BCI): one feature based on multifractal cumulants, and one feature based on the predictive complexity of the EEG time series. The multifractal cumulants feature measures the signal regularity, while the predictive complexity measures the difficulty to predict the future of the signal based on its past, hence a degree of how complex it is. We have conducted an evaluation of the performance of these two novel features on EEG data corresponding to motor-imagery. We also compared them to the most successful features used in the BCI field, namely the Band-Power features. We evaluated these three kinds of features and their combinations on EEG signals from 13 subjects. Results obtained show that our novel features can lead to BCI designs with improved classification performance, notably when using and combining the three kinds of feature (band-power, multifractal cumulants, predictive complexity) together.

## 1 Introduction

Brain-Computer Interfaces (BCI) are communication systems that enable users to send commands to a computer by using only their brain activity [27], this activity being generally measured using ElectroEncephaloGraphy (EEG). Most EEG-based BCI are designed around a pattern recognition approach: In a first step features describing the relevant information embedded in the EEG signals are extracted [4]. They are then feed into a classifier which identifies the class of the mental state from these features [17]. Therefore, the efficiency of a BCI, in terms of recognition rate, depends mostly on the choice of appropriate features and classifiers. Despite the large number of features that have been explored to design BCI [4], the performances of current EEG-based BCI are still not satisfactory, and the BCI community has stressed the need to further explore and design alternative features [19].

In this paper, we focus on feature extraction from EEG signals for the design of BCI based on motor imagery (MI) [22]. MI corresponds to the imagination of limb movements (e.g., hand or feet) without actual output. It is known to trigger brain signals variations in specific frequency bands that can be used to drive a BCI [22]. The contribution of this paper is two-fold. First, we introduce two new features for MI classification in EEG-based BCI. These two new features are based on: 1. multifractal cumulants [26] and 2. predictive complexity [9]. Second, we perform systematic comparisons and analysis of the performance of these two new features, together with the most successful feature for motor-imagery based-BCI, namely, band-power feature [13].

The first new feature, namely, multifractal cumulants, can be seen as a statistic on inter-frequency band relations. This is particularly relevant for BCI as this information is generally ignored in current motor imagery-based BCI designs, mostly based on the sole power in different frequency bands. It should be mentioned that a preliminary study of this kind of feature has been presented in a conference paper [8]. The second new feature is based on the statistical complexity and predictive properties of the time

---

\*Nicolas is with the Laboratoire du Traitement du Signal et de l'Image (LTSI) at University of Rennes 1, Campus de Beaulieu, Bât 22, 35042 Rennes Cedex - France.

†Fabien is with the Institute for Infocomm Research (I2R), Signal Processing Department, Brain-Computer Interface Laboratory, at 1 Fusionopolis way, 138632, Singapore.

‡Anatole is with the Institut National de Recherche en Informatique et en Automatique (INRIA), at Campus universitaire de Beaulieu, Avenue du Général Leclerc, 35042 Rennes Cedex - France.

series [9]. The information (quantified in bits) that is extracted this way measures how difficult it is to make an optimal prediction based on past information. It is null both for totally ordered and totally random systems, and increases in between. It has already been applied to single simulated neurons [7] and a related form was applied to measure synchronisation in the brain [15]. The assumption is that performing a mental task (e.g., motor imagery) makes the EEG signal either more or less predictable, which can be detected by a classifier when quantified by this second new feature.

The remainder of this paper is organised as follows: Sections 2 and 3 present the two new features that we propose. Section 4 presents the experimental evaluation, including the data sets that were used. The results are then discussed. Section 5 concludes this study.

The code used for all the experiments in this paper is provided as Free/Libre software. The data used is available online and all the presented results are reproducible independently.

## 2 Multifractal cumulants

The multifractal formalism is described in details in [26, 21]. This section presents a short overview for the needs of this document.

Intuitively, the multifractal cumulants of the signal capture a signature of inter-band relations (see below). This contrasts to the power in each frequency band that is generally used. As shown in [8] the multifractal spectrum can in itself be used for EEG classification. When considering multifractal in addition to power band feature vectors, the resulting combination may improve the classification accuracy.

The method we chose for extracting the multifractal spectrum is a discrete wavelet transform of the signal, out of which we extract the wavelet leader coefficients [2]. Following the directions of [26] we then use the cumulants of the leaders as the features for classification, unlike what we previously did in [8].

Let  $x(t)$  be the signal to analyse. One view on multifractal analysis [18] is to relate the statistical properties of  $x(t)$  and of a scaled version of it  $x(at)$ . In terms of frequency analysis, that scaling in time corresponds to a frequency shift. Hence, another view of the multifractal cumulants feature is that they characterise some form of inter-frequency information, as mentioned in the introduction of this section.

More precisely:

- The signal  $x(t)$  is decomposed using a Discrete Wavelet Transform to get the wavelet coefficients  $w(s, t_s)$  at each dyadic scale  $s$  and time interval  $t_s$ .
- The wavelet leaders at each scale  $s$  are then extracted by computing the maxima of the wavelet coefficients over all samples involved for computing  $w(s, t_s - 1)$ ,  $w(s, t_s)$  and  $w(s, t_s + 1)$  (including lower scales) [2].
- Instead of performing a Legendre transform, or a direct Holder exponent density estimation as in [8], we use here the recent technique introduced by [26] and compute the wavelet leader cumulants of orders 1 to 5. As noted in [26] the first few cumulants already contain most of the information useful in practice for characterising the distribution of the Holder exponents. For a classification task this information can now be exploited in a more condensed form.
- The 5 first cumulants are computed for the leaders at each scale  $s$ . Considering there is at most  $L$  levels of wavelet transform in a signal of size between  $2^L$  and  $2^{L+1}$ , we get a total of  $5 \times L$  cumulants for the signal, that progressively encompass more and more frequency bands as the scale increases. These  $5 \times L$  cumulants per channel are used as the feature vector.

This method can be quite sensitive to the presence of electromagnetic interferences at 50 Hz. We thus pre-filter the signals as described in section 4.1 before proceeding to the multifractal cumulants estimation.

## 3 Predictive complexity measure

This paper introduces for the first time the predictive complexity measure of [9] in the context of EEG classification.

The intuitive idea behind this feature is to quantify the amount of information that is necessary to retain from the past of the series in order to be able to predict optimally the future of the series ([24], and see below for the optimality criterion).

We had indication from related previous works that the feature could be relevant for EEGs:

- At the level of neurons: statistical complexity was used to describe the computational structure of spike trains [12]. It was also shown to decrease while the neurons are learning in an artificial spiking neural network [7].
- At the level of the brain: information coherence and synchronisation mechanism between communities of neurons are presented in [15], relying on related techniques.

### 3.1 Decisional states and the corresponding complexity measure

Informally, the idea behind the decisional states is to construct a Markovian automaton [11, 24] whose states correspond to taking the same decisions [9], according to a user-defined utility function. These decisions are those that one can take based on predictions of the future and their expected utility.

The complexity of the series is then computed as the mutual information between the internal states of the Markovian automaton, and the series itself. The complexity is null for a very regular series, for example a constant series or a series where we always take the same decision: there is only a single state in the automaton. Similarly the complexity of a completely random series is also null: it can be modelled by successive independent draws from a fixed probability distribution, whose expected utility we take to make our decision. This leads again to a single Markovian automaton state, hence a null complexity. The complexity measure increases only for more complicated series with many internal states (i.e. many distinct probability distributions of what happens next, depending on what previously happened, leading to different decisions).

Presumably when the EEG corresponds or not to some functional activity, the complexity of the series should change. The idea is to plug machine learning techniques for monitoring that change.

#### Formal description

Formally, let  $(s_t)$  be a time series, with  $t$  the time index. Let  $s_t^{-\infty} = (s_u)_{u \leq t}$  and  $s_t^{+\infty} = (s_u)_{t < u}$  be the past and future histories at time  $t$ . In practice and for real measures, the time range is finite:  $0 \leq t \leq T$ . Similarly we measure the past and future histories with finite horizons:  $s_t^{-h} = (s_u)_{t-h \leq u \leq t}$  and  $s_t^{+k} = (s_u)_{t < u \leq t+k}$ .

Let us now consider predicting the future from the past statistically: we seek to determine  $P(s_t^{+k} | s_t^{-h})$  at each time  $t$  (possibly with  $h$  or  $k$  infinite). Assuming the system is conditionally stationary, that distribution does not change through time: the same causes produce the same consequences. Let then  $S^{-h} = \{s_t^{-h}\}_{\forall t}$  and  $S^{+k} = \{s_t^{+k}\}_{\forall t}$  the sets of all past and future histories. We will drop the time indices from now on to indicate the time shift invariance.

The causal states  $\zeta$  are defined as the equivalence classes of past histories with the same conditional distribution of futures:  $\zeta(s^{-h}) = \{x \in S^{-h} : P(s^{+k} | x) = P(s^{+k} | s^{-h})\} = \{x : x \stackrel{c}{\equiv} s^{-h}\}$ . Knowing the causal state at the current time is the minimal information needed for making optimal predictions [24] using the full conditional probability distribution.

In the discrete case the series are strings of symbols drawn from an alphabet  $\mathcal{A}$ . Each time step implies a symbol transition, which possibly leads to a different causal state. The corresponding automaton is called the  $\epsilon$ -machine [11].

Let us now introduce a utility function  $u : (S^{+k})^2 \mapsto \mathbb{R}$ , such that  $u(r^{+k}, s^{+k})$  quantifies the gain (positive) or loss (negative) when the user relied on the prediction  $r^{+k}$  while  $s^{+k}$  actually happened. We can now define an expected utility:  $\mathbb{U}[r^{+k} | s^{-h}] = \mathbb{E}_{s^{+k} \in S^{+k}} [u(r^{+k}, s^{+k}) | s^{-h}]$ , quantifying what utility can be expected on average when choosing the prediction  $r^{+k}$  for the current system state  $s^{-h}$ . The set of optimal predictions, realising the maximal expected utility, can now be defined as  $Y(s^{-h}) = \text{argmax}_{r^{+k} \in S^{+k}} \mathbb{U}[r^{+k} | s^{-h}]$ .

The following equivalence relations  $\stackrel{p}{\equiv}$ ,  $\stackrel{u}{\equiv}$ , and  $\stackrel{d}{\equiv}$  naturally extend the causal state equivalence relation  $\stackrel{c}{\equiv}$ , taking into account the utility function:

- $r^{-h} \stackrel{p}{\equiv} s^{-h}$  when  $Y(r^{-h}) = Y(s^{-h})$ , with the corresponding *iso-prediction sets* as equivalence classes. All past histories within the same class lead to choosing the same predictions, even though the expected utility may change from one past history to the other.
- $r^{-h} \stackrel{u}{\equiv} s^{-h}$  when  $\max_{r^{+k} \in S^{+k}} \mathbb{U}[r^{+k} | r^{-h}] = \max_{r^{+k} \in S^{+k}} \mathbb{U}[r^{+k} | s^{-h}]$ , with the corresponding *iso-utility sets* as equivalence classes. All past histories within the same class lead to the same maximum

expected utility, even though the optimal predictions to choose for reaching this utility are not specified.

- $r^{-h} \stackrel{d}{\equiv} s^{-h}$  when  $r^{-h} \stackrel{p}{\equiv} s^{-h}$  and  $r^{-h} \stackrel{u}{\equiv} s^{-h}$ . The intersection of the above iso-utility and iso-prediction sets define the *decisional states*:  $\Psi(s^{-h}) = \{x \in S^{-h} : x \stackrel{p}{\equiv} s^{-h}, x \stackrel{u}{\equiv} s^{-h}\}$ . We assume that when both the maximal expected utility and the optimal predictions are the same, the user will reach the same decisions. In other words, the utility function encodes all the user needs to know to reach a decision.

It can be easily shown [9] that the causal states sub-partition the decisional states. That is to say, the causal states have lost their minimality property due to the fact we are not interested in the full conditional distribution of futures but only in the optimal decisions with respect to a user-defined utility function.

The causal states are an intrinsic property of the data set. The mutual information  $M_c = I(s^{-h}; \zeta(s^{-h}))$  between the causal  $\zeta(s^{-h})$  state and the series of  $s^{-h}$  defines an intrinsic measure, the statistical complexity [11], quantifying how hard it is to get the conditional distribution of futures from the current observed past.

The decisional states correspond to the structure implied by the user utility function on top of the causal states. As for the causal states, knowing the decisional state at the current time is the information needed for making optimal predictions maximising the user-defined utility function. The mutual information  $M_d = I(s^{-h}; \Psi(s^{-h}))$  similarly defines a complexity measure for how hard it is to make these predictions, called decisional complexity by analogy with the statistical complexity.

### 3.2 Application to EEG data

In the present study the chosen utility function is the negative sum of square error:  $u(r^{+k}, s^{+k}) = -\|r^{+k} - s^{+k}\|^2$ .

Each EEG series is split in time windows of  $h + k$  samples, defining each a  $(s^{-h}, s^{+k})$  pair. These observations are fed in the reconstruction algorithm presented in [9]. That algorithm estimates the joint probabilities  $p(s^{-h}, s^{+k})$  using a kernel density estimation with Gaussian kernels. The conditional probabilities are then computed by integration over  $S^{+k}$ :  $p(s^{+k}|s^{-h}) = p(s^{-h}, s^{+k}) / \int_{\sigma^{+k} \in S^{+k}} p(s^{-h}, \sigma^{+k})$ . These are then clustered into causal state estimates  $\zeta(s^{-h}) = \{x \in S^{-h} : p(s^{+k}|x) = p(s^{+k}|s^{-h})\}$  using connected components up to a fixed Bhattacharyya distance threshold (chosen to be 0.05) between the conditional distributions. The utility function is applied on top of the aggregated causal states distributions  $p(s^{+k}|\zeta) = \text{avg}_{s^{-h} \in \zeta} p(s^{+k}|s^{-h})$  in order to get the expected utilities  $\mathbb{U}[r^{+k}|\zeta]$ . Finally, the causal states are themselves clustered into iso-utility and iso-prediction sets, which are intersected to get the decisional states  $\omega$ .

The decisional complexity is the feature that is extracted from the EEG series. We thus get one feature per electrode.

## 4 Evaluation

### 4.1 Data sets

Four data sets for a total of 13 subjects were used for evaluation in this study. Data for 12 subjects come from the international BCI competitions<sup>1</sup> II, III, and IV [5][6]. The data for the last subject was acquired at INRIA (French National Research Institute on Computer Science and Control) Rennes-Bretagne Atlantique, using the OpenViBE software platform [23]<sup>2</sup>.

#### 4.1.1 BCI competition II, data set III (1 subject)

This data set was captured at the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz [1]. The data contains 280 trials sampled at 128 Hz. During each trial, the subject is presented with a visual cue indicating either left or right at random, and shall then imagine a movement of the corresponding hand during 6s. There was 140 trials of left class (left hand motor imagery) and 140 trials of the right class (right hand motor imagery). EEG were recorded using

<sup>1</sup><http://www.bbc.de/competition/>

<sup>2</sup><http://openvibe.inria.fr/?q=datasets>

the C3, C4 and Cz electrodes, however, for the purpose of this evaluation, we used only the C3 and C4 electrodes as recommended in [16]. More details about this data set can be found in [5].

The data was already preprocessed by a band-pass filter between 0.5 and 30 Hz. Unfortunately, the nature of the filter is not specified, and the DC component was not well removed. Since this interferes in particular with the algorithm for estimating the series complexity, a new filter is applied with the following characteristics:

- FIR filter obtained using METEOR [25]
- Less than 1dB change in 4-30Hz with a linear phase response
- At least -50dB at 0 Hz (and above 40 Hz, though in the present case the signal is already filtered)
- 1/4 sec delay

We then used all the available filtered data over the feedback period in order to extract the features.

#### 4.1.2 BCI competition III, data set IIIb (2 subjects)

This data set was also captured at the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz [1]. It originally consists of EEG signals from 3 subjects who performed left hand and right hand motor imagery. However, for the purpose of this study, only subjects labeled S4 and X11 were used. Indeed, EEG signals for subject O3VR were recorded using a different protocol and the data file provided online contained erroneously duplicate signals<sup>3</sup>. The experimental protocol for subjects S4 and X11 is similar to the one used for the BCI competition II data set III. The data was captured at a 125 Hz sampling rate using electrodes C3 and C4. More details about this data set can be found in [6]. EEG signals in this data set were already band-pass filtered in 0.5-30 Hz. As for the previous data set, an additional FIR filter with the same characteristics as the previous one was applied, for the same reasons. We also used all the available filtered data over the feedback period in order to extract the features.

#### 4.1.3 BCI competition IV, data set I Ib (9 subjects)

This third data set was provided by the same team and follows the same experimental protocol as the above two. However it contains ocular artifacts that interfere with the brain signals. Although this is less convenient for the purpose of testing our two new features it is also more realistic. We thus choosed to ignore the artifacts and process as if the signals were clean EEGs, in order to assert the robustness of our features to the ocular artifacts.

Data coming from nine subjects is sampled at 250 Hz, and pre-processed by a band-pass filter between 0.5 and 100 Hz with a notch at 50 Hz. For the aforementioned reasons we had to re-filter the signals. We choosed a FIR design with a band-pass between 6 and 30 Hz, with a null at 0Hz to suppress the DC component and -50dB above 40 Hz.

We also used all the available filtered data over the feedback period in order to extract the features.

#### 4.1.4 OpenViBE / INRIA data (1 subject)

This data set was captured at INRIA Rennes-Bretagne Atlantique using the OpenViBE free and open-source software [23]. This data set comprises EEG signals from one subject who performed left hand and right hand motor imagery. 560 trials of motor imagery (280 trials per class) were recorded over a 2 week period. Data were collected using the same experimental protocol as the one used for the BCI competition data, i.e., the Graz BCI protocol [22]. Half of the trials (randomly selected from all experiments over time) is used for training, the remaining half for testing.

EEG data was sampled at 512 Hz and recorded using the following electrodes: C3, C4, FC3, FC4, C5, C1, C2, C6, CP3, CP4, with a nose reference electrode. The use of such electrodes enables us to apply a discrete Laplacian spatial filter [14] over C3 and C4 in order to obtain better signals, as recommended in [20].

The data was preprocessed by a FIR filter (designed with METEOR [25]) with the following characteristics:

- Less than 1dB change in 4-30Hz, linear phase response in this range.

---

<sup>3</sup>See [http://www.bbci.de/competition/iii/desc\\_IIIb\\_ps.html](http://www.bbci.de/competition/iii/desc_IIIb_ps.html) for details

- At least -50dB at 0 Hz and above 50 Hz.
- Null responses at 50 Hz and all harmonics.
- 1/4 sec delay.

We then used all the available filtered data over the feedback period in order to extract the features.

## 4.2 Results obtained with all features and their combinations

The goal of this section is to show what results can be obtained with each feature considered in isolation, and the effect of combining them. The hypothesis we wish to test is that each feature extracts a different information from the signal, and thus that combining them can improve the classification accuracy.

### 4.2.1 Features

For the experiments below, the following methodology was applied:

- The power in frequency bands was extracted for each subject. As was shown by [13] and in our own forthcoming study [10] estimating this information from data is sensitive to choice of the time-frequency decomposition algorithm. We extracted the power information in each frequency band using a Morlet Wavelet which support is determined by cross-validation. All bands were then kept between 4 and 30 hertz, for each of the C3 and C4 electrode: this leads to a 52-dimensional feature vector.
- The multifractal cumulants feature was extracted according to the method described in Section 2. Cross-validation on the training set was used to determine the wavelet support and the number of decomposition levels to retain. Since we have 5 cumulants per level per electrode, this leads to a  $10 * N$ -dimensional vector with  $N$  the number of retained levels.
- Parameters for estimating the predictive complexity feature were also determined by maximising the cross-validation performance on the training set: the number of points to retain from the past and the future, the sub-sampling factor for the series, the kernel size used for estimating the conditional probability distributions, and whether to work on the raw series or the first differences. Cross-validation selected 1 point in the future, 5 points from the past with a sub-sampling factor of 8 for the OpenViBE subject, and 6 points from the past with a sub-sampling factor of 2 for the BCI II and III subjects. These sub-sampling parameters correspond to having a sampling frequency of 62.5 Hz for the S4 and X11 subjects and 64 Hz for the others, which thanks to Nyquist theorem matches the filtering operation that was performed on the signals: cross-validation selected the most economical sub-sampling parameter that still captures the remaining frequencies in the signal between 4 and 30 Hz. We thus saved computational time for the BCI IV subjects by fixing the subsampling parameters and cross-validating only the kernel width. All sub-sampled signals (ex: all 8 possible series for a sub-sampling by factor 8) were kept for building the statistics in the complexity computations.

Therefore the feature vectors include:

- 52 power features for frequencies between 4 and 30 Hz. (26 for each C3 / C4 electrode)
- 30, 40, or 50 features for multifractal cumulants.
- 2 features for the predictive complexity (one for each electrode).

As a side remark we shall point out that cross-validating all the above parameters required a quite consequent computational power, which was obtained using the BOINC distributed computing infrastructure [3]. Nevertheless, once the optimal parameters were determined off-line by cross-validation, the computational requirements for extracting the features are nothing exceptional: the features can be computed in real time on a standard PC.

Subjet	Band-Power	Multifractal	Complexity	Combination	Gain
bci2	77.1	80.7	77.9	80.7	3.57
s4	81.5	74.4	65.0	81.7	0.19
x11	80.4	68.7	70.2	80.9	0.56
OpenVibe	92.9	85.7	76.1	93.2	0.36
bci4s1	77.5	65.9	55.9	74.4	-3.12
bci4s2	56.4	57.5	56.1	56.1	-0.36
bci4s3	51.9	51.2	54.1	52.2	0.31
bci4s4	93.4	91.2	90.3	95.6	2.19
bci4s5	96.9	88.8	74.1	95.9	-0.94
bci4s6	87.8	82.2	68.4	89.1	1.25
bci4s7	70.6	68.1	70.3	72.2	1.56
bci4s8	80.0	88.4	90.6	89.1	9.06
bci4s9	78.8	83.1	78.8	82.8	4.06

Legend: Green when the classification accuracy of the combination is higher by at least 1% than when using only the usual Band-Power feature, blue when the change in performance with the combination is between -1% and 1%, orange when the combination is worse.

Table 1: Classification accuracies over the test sets for each subject and feature

#### 4.2.2 Classifier and combination rule

In order to classify the extracted features, we used a Linear Discriminant Analysis (LDA), one of the most popular and efficient classifier for EEG-based BCI [17].

Each series was classified independently into either the “left hand” or “right hand” class of the motor imagery task using each feature independently. The percentage of correct classifications on the test set is reported for each sujet in the first 3 colums of Table 1.

We also evaluated the accuracy obtained when combining these three features together. To do so, we first estimate the global accuracy of each feature alone by using the Fisher discriminant ratio  $F$  on the training data set, hence we get  $F^{\text{pow}}$ ,  $F^{\text{mfa}}$ ,  $F^{\text{cpx}}$  for the three features Band-Power, Multifractal cumulants, and Complexity. In order to combine the results we gather the outputs of the 3 linear classifiers trained independently on each feature by performing a simple arithmetic average for each instance  $P_i^{\text{combi}} = F^{\text{pow}} * P_i^{\text{pow}} + F^{\text{mfa}} * P_i^{\text{mfa}} + F^{\text{cpx}} * P_i^{\text{cpx}}$  where each  $P_i$  is a prediction (classifier output) for the instance number  $i$ . That value is then used for classifying the instance in the results reported in the fourth column of table 1.

#### 4.2.3 Results

The multifractal cumulants and the predictive complexity features allow by themselves to classify the BCI data sets with a good accuracy. This, in itself, is a contribution of this paper. Although the accuracy obtained using either the Multifractal Cumulants or the Predictive Complexity is slightly lower on average that when using only the Power Bands feature (sometimes higher), we have extracted some information from the signal that is adequate for classification.

Results also showed that combining all features is better than using the Power Bands alone for 6 subjects, on the same order for 6 subjects, and with a negative impact for only one subject. A gain of 1.4% accuracy was obtained on average, with much higher values for some subjects (see table 1). This not only confirms that some different information was extracted from the signal, but also stresses the usefulness of our new features for BCI classification tasks.

## 5 Conclusion

This study has introduced two new features for Brain-Computer Interface design: multifractal cumulants and predictive complexity. The information contained in the multifractal cumulants feature corresponds to a relation between frequency bands, rather than the power extracted in each band. The complexity feature measures the difficulty to predict the future of the EEG signals based on their past.

Interestingly enough, our results showed that the two new features, i.e., multifractal cumulants and predictive complexity measure, could indeed be used by themselves to discriminate between different motor imagery mental states as measured by EEG. This is especially interesting as the complexity feature

only adds one scalar value per electrode, compared to the other features we considered (ex: one dimension per frequency band in the band-power case).

Moreover, our results showed that when combining these two features together with band-power features, the resulting BCI could reach a better classification accuracy than when using the usual band-power feature alone. Thus this suggests that these new features are a good complement to currently used features for BCI design and that they can lead to improved BCI design. Therefore we would recommend BCI designers to consider these two features as additional features in the conception of a BCI system, in order to obtain better performance.

Future work might involve the exploration of novel ways to combine the features and feature selection, as well as the application of these features to BCI tasks other than motor imagery. Work is also needed on the design of novel algorithms including physiologically relevant error functions for EEG signal predictions for the complexity feature.

## Appendix: Source code, data, and web information

All the results in this study are reproducible independently. The code for the experiments is provided as Free/Libre software on the main author web site:

- <http://nicolas.brodu.numerimoire.net/en/recherche/publications/>

The data that was used can be downloaded on the BCI competitions website and on the OpenViBE project page:

- <http://www.bbci.de/competition/>
- <http://openvibe.inria.fr/?q=datasets>

## References

- [1] Schlögl A., Neuper C., and Pfurtscheller G. Estimating the mutual information of an eeg-based brain-computer-interface. *Biomedizinische Technik*, 47(1-2):3–8, 2002.
- [2] Patrice Abry, Stéphane Jaffard, and Bruno Lashermes. Revisiting scaling, multifractal, and multiplicative cascades with the wavelet leader lens. In *SPIE*, pages 103–117, 2004.
- [3] David P. Anderson. Boinc: A system for public-resource computing and storage. In *5th IEEE/ACM International Workshop on Grid Computing.*, November 2004.
- [4] A. Bashashati, M. Fatourech, R. K. Ward, and G. E. Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2):R35–57, 2007.
- [5] B. Blankertz, K. R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer. The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. *IEEE Transactions on Biomedical Engineering*, 51(6):1044–1051, 2004.
- [6] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. D. R. Millan, M. Schroder, and N. Birbaumer. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):153–159, 2006.
- [7] Nicolas Brodu. Quantifying the effect of learning on recurrent spikin neurons. In *IJCNN*, pages 512–517, 2007.
- [8] Nicolas Brodu. Multifractal feature vectors for brain-computer interfaces. In *IJCNN*, pages 2883–2890, 2008.
- [9] Nicolas Brodu. Decisional states. <http://arxiv.org/abs/0902.0600>, submitted, 2009.

- [10] Nicolas Brodu, Fabien Lotte, and Anatole Lécuyer. Comparative study of band-power extraction techniques for eeg-based brain-computer interfaces. submitted.
- [11] James Crutchfield and Karl Young. Inferring statistical complexity. *Physical Review Letters*, 62(2):105–108, 1989.
- [12] Rob Haslinger, Kristina Klinkner, and Cosma Rohilla Shalizi. The computational structure of spike trains. *Neural Computation*, 22:121–157, 2010.
- [13] Pawel Herman, Girijesh Prasad, Thomas Martin McGinnity, and Damien Coyle. Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(4), August 2008.
- [14] B. Hjorth. An on-line transformation of eeg scalp potentials into orthogonal source derivations. *Electroencephalography and Clinical Neurophysiology*, 39:526–530, 1975.
- [15] Kristina Klinkner, Cosma Rohilla Shalizi, and Marcelo Camperi. Measuring shared information and coordinated activity in neuronal networks. *Advances in Neural Information Processing Systems*, 18:667–674, 2006.
- [16] S. Lemm, C. Schafer, and G. Curio. BCI competition 2003–data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements. *IEEE Transactions on Biomedical Engineering*, 51(6):1077–1080, 2004.
- [17] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 4:R1–R13, 2007.
- [18] Benoit Mandelbrot, Adlai Fisher, and Laurent Calvet. A multifractal model of asset returns. Technical Report 1164, Cowles Foundation, 1997.
- [19] D. J. McFarland, C. W. Anderson, K.-R. Muller, A. Schlogl, and D. J. Krusienski. BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):135–138, 2006.
- [20] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalographic Clinical Neurophysiology*, 103(3):386–394, 1997.
- [21] Jean-François Muzy, Emmanuel Bacry, and Alain Arneodo. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E.*, 47:875–884, 1993.
- [22] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *proceedings of the IEEE*, 89(7):1123–1134, 2001.
- [23] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer. Openvibe: An open-source software platform to design, test and use brain-computer interfaces in real and virtual environments. *Presence: teleoperators and virtual environments*, 19(1), 2010.
- [24] Cosma Rohilla Shalizi. *Causal Architecture, Complexity, and Self-Organization in Time Series and Cellular Automata*. PhD thesis, University of Wisconsin at Madison, 2001.
- [25] K. Steiglitz, T. W. Parks, and J. F. Kaiser. Meteor: A constraint-based fir filter design program. *IEEE Transactions on Signal Processing*, 40(8):1901–1909, 1992.
- [26] Herwig Wendt, Patrice Abry, and Stéphane Jaffard. Bootstrap for empirical multifractal analysis. *IEEE Signal Processing Magazine*, (38), 2007.
- [27] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.